

# Maximum likelihood estimation for stationary point processes

(compensator/stochastic intensity/martingale/natural increasing process/point process/predictable process)

MADAN L. PURI<sup>†</sup> AND PHAM D. TUAN<sup>‡</sup>

<sup>†</sup>Department of Mathematics, Indiana University, Bloomington, IN 47405; and <sup>‡</sup>Mathématiques Appliquées, Université de Grenoble, 38041 Grenoble Cedex, France

Communicated by Herbert Robbins, September 27, 1985

**ABSTRACT** In this paper we derive the log likelihood function for point processes in terms of their stochastic intensities by using the martingale approach. For practical purposes we work with an approximate log likelihood function that is shown to possess the usual asymptotic properties of a log likelihood function. The resulting estimates are strongly consistent and asymptotically normal (under some regularity conditions). As a by-product, a strong law of large numbers and a central limit theorem for martingales in continuous times are derived.

## The Log Likelihood Function for Point Processes

We shall be concerned with multitype point processes, that is, point events of  $r$  different types randomly occurring along the real line. This process can be described by a multivariate counting process  $N(t) = \{N_1(t), \dots, N_r(t)\}$ ,  $t \in R$ , defined on some probability space  $(\Omega, \mathcal{A}, P)$ . Here  $N_j(t) - N_j(s)$ ,  $t > s$  denotes the number of point events of type  $j$  that occur in  $(s, t]$ . By convention  $N_j(0) = 0$ . We shall suppose that at each  $t$ , at most one event, regardless of its type, can occur. Let  $\mathcal{A}_t$ ,  $t \geq 0$  be an increasing family of sub- $\sigma$  fields such that  $N(t)$  is  $\mathcal{A}_t$ -measurable,  $t > 0$ . Then there is for each  $j = 1, \dots, r$  a natural increasing process  $\Lambda_j(t)$ , called the compensator of  $N_j(t)$ , relative to  $(\mathcal{A}_t, P)$  such that  $m_j(t) = N_j(t) - \Lambda_j(t)$ ,  $t \geq 0$  is an  $\mathcal{A}_t$ -local martingale. Here we shall be interested only in the case in which the measure  $d\Lambda_j(t)dP$  admits a density  $\lambda_j(t)$  with respect to  $dt dP$ . The processes  $\lambda_j(t)$  can be chosen to be  $\mathcal{A}_t$ -predictable; that is,  $(t, \omega) \rightarrow \lambda_j(t, \omega)$  is measurable with respect to the  $\sigma$ -field generated by all the  $\lambda_r$ -adapted processes with left continuous sample paths and is called the stochastic intensity of the  $N(t)$  process. Intuitively,  $\lambda_j(t)$  can be interpreted as

$$\lim_{\Delta t \downarrow 0} P\{\Delta N_j(t) = 1 | \mathcal{A}_t\} / \Delta t, \quad [1]$$

where  $\Delta N(t) = N(t + \Delta t) - N(t)$ . Indeed, if  $\tau_n^j$  is the time of occurrence of the  $n$ th event of type  $j$  after the origin, then  $m_j(t \wedge \tau_n^j)$ ,  $t \geq 0$  is a martingale and so for  $s > t$ ,  $A \in \mathcal{A}_t$ ,

$$\begin{aligned} E[1_A \{N_j(s \wedge \tau_n^j) - N_j(t \wedge \tau_n^j)\}] &= E\left\{1_A \int_{t \wedge \tau_n^j}^{s \wedge \tau_n^j} d\Lambda_j(u)\right\} \\ &= \int_t^s E[1_A 1_{\{u \leq \tau_n^j\}} \lambda_j(u)] du. \quad [2] \end{aligned}$$

Heuristically, if  $s = t + \Delta t$  and  $\Delta t$  is small, then the left-hand side of Eq. 2 is approximately  $P(A \cap \{t \leq \tau_n^j\} \cap \{\Delta N(t) = 1\})$  and the right-hand side of Eq. 2 is approximately  $E[1_A 1_{\{t \leq \tau_n^j\}} \lambda_j(t)] \Delta t$ . Since  $\{\tau_n^j < t\} \in \mathcal{A}_t$ , we obtain the interpretation 1. Historically Eq. 1 was proposed as the definition of the sto-

chastic intensity. This definition requires the existence of the limit in Eq. 1 and is equivalent to our definition only under some regularity conditions (1).

If  $\mathcal{A}_t = \tilde{\mathcal{A}}_t$  represents the sub  $\sigma$ -field generated by  $N(s)$ ,  $0 \leq s \leq t$ , then the stochastic intensity, which we now write as  $\lambda_j(t)$ , completely defines the probability distribution of the process (2, 3). Let  $\pi$  be the probability measure under which the  $N_j(t)$  are independent Poisson processes with unit rate. Observe that the random measure  $\{j\} \times (s, t) \rightarrow \int_s^t \lambda_j(u) du$  on  $\{1, \dots, r\} \times R^+$  is precisely the dual predictable projection of the random measure  $\{j\} \times (s, t) \rightarrow N_j(t) - N_j(s)$  in the sense of Jacod (2). By a result of this paper,  $P$  is absolutely continuous with respect to  $\pi$  on any  $\tilde{\mathcal{A}}_T$ ,  $T > 0$ , with density

$$\frac{dP}{d\pi} = \exp\left(\int_0^T \sum_{j=1}^r \{\log \tilde{\lambda}_j(t) dN_j(t) - \tilde{\lambda}_j(t) dt\}\right).$$

Now let  $\{P_\theta, \theta \in \Theta\}$  be a family of probability distributions on  $(\Omega, \mathcal{A})$  and let  $\lambda_{\theta,j}(t)$  be the corresponding stochastic intensity of the  $N_j(t)$  process. Then our result shows that the log likelihood function corresponding to an interval of observation  $[0, T]$  is, up to an additive constant,

$$\tilde{L}_T(\theta) = \int_0^T \sum_{j=1}^r \{\log \tilde{\lambda}_{\theta,j}(t) dN_j(t) - \tilde{\lambda}_{\theta,j}(t) dt\}. \quad [3]$$

**Remark:** The multitype process is a special case of the marked point process when the space of the marks is just  $\{1, \dots, r\}$ . It is not necessary to suppose the existence of the stochastic intensity to write down the log likelihood function for the marked point process; it suffices that the measure  $d\tilde{\Lambda}_{\theta,j}(t)dP_\theta$  be absolutely continuous with respect to some measure  $d\nu(t)dP_\theta$ , where  $\tilde{\Lambda}_{\theta,j}(t)$  is the natural increasing process of  $N_j(t)$  relative to  $\tilde{\mathcal{A}}_t$  and to  $P_\theta$ . The special case we considered is convenient for further development concerning the asymptotic properties of the log likelihood function.

## The Approximate Log Likelihood Function

The process we consider is defined on the whole line although only an observation on  $[0, T]$  is available. Denote by  $\lambda_{\theta,j}(t)$  the stochastic intensity of the  $N_j(t)$  process relative to  $P_\theta$  and the sub  $\sigma$ -fields  $\mathcal{A}_t$ ,  $t \in R$  generated by  $N(s)$ ,  $-\infty < s < t$ . Models of point processes are usually described in terms of  $\lambda_\theta(t) = (\lambda_{\theta,1}(t), \dots, \lambda_{\theta,r}(t))$ . For example, the self-exciting process introduced by Hawkes (4) and also considered by Hawkes and Oakes (5) can be defined by

$$\lambda_\theta(t) = \alpha_\theta + \int_{-\infty}^t g_\theta(t-s) dN(s), \quad [4]$$

where  $\alpha_\theta$  is a constant vector and  $g_\theta(\cdot)$  is some appropriate matrix function. Thus, it is desirable to obtain the log likelihood function in terms of  $\lambda_\theta(t)$ . We are led to the problem of computing  $\tilde{\lambda}_\theta(t)$  in terms of  $\lambda_{\theta,j}(t)$ . Now, from the interpretation 1 of  $\lambda_{\theta,j}(t)$  one can expect that  $\tilde{\lambda}_{\theta,j}(t) = E_\theta$

$\{\lambda_{\theta,j}(t)|\mathcal{A}_t\}$  in case  $\lambda_{\theta,j}(t)$  is integrable. The following theorem and its corollary are of the type often called *innovation theorems* (see, e.g., p. 372 of ref. 6).

**THEOREM 1.** Let  $\tau_n^1$  be the time of occurrence of the  $n$ th event of type  $j$  after the origin. Then, for almost all  $t$ , the random variable  $1_{\{t \leq \tau_n^1\}} \lambda_{\theta,j}(t)$  is  $\mathcal{P}_\theta$ -integrable and  $1_{\{t \leq \tau_n^1\}} \lambda_{\theta,j}(t) = E_\theta 1_{\{t \leq \tau_n^1\}} \lambda_{\theta,j}(t) | \mathcal{A}_t$  almost surely.

*Proof:* From Eq. 2, we have for all  $s > 0$ ,

$$n \geq \int_0^s E_\theta [1_{\{t \leq \tau_n^1\}} \lambda_{\theta,j}(t)] dt$$

and hence the function

$$(t, \omega) \rightarrow X_n(t, \omega) = 1_{\{t \leq \tau_n^1\}}(\omega) \lambda_{\theta,j}(t)$$

belongs to  $L^1(R^+ \times \Omega, \mathcal{B}_R \otimes \mathcal{A}, dtdP_\theta)$ . We show that there exists an  $\mathcal{A}_t$ -predictable process  $\tilde{X}_n(t)$  with  $\tilde{X}_n(t) = E\{X_n(t)|\mathcal{A}_t\}$  for almost all  $t$ , almost surely. Indeed, there exists a sequence of simple functions of the form

$$X_{n,k}(t, \omega) = \sum_{m=1}^{m_k} Z_m^{(k)}(\omega) 1_{(\tau_{m-1}^{(k)}, \tau_m^{(k)}]}(t), \quad k = 1, 2, \dots,$$

which converges in  $L^1$  to  $X_n$ . For  $t_{m-1}^{(k)} < t \leq t_m^{(k)}$ , set  $\tilde{X}_{n,k}(t)$  to be a version of  $E_\theta\{Z_m^{(k)}|\mathcal{A}_t\}$ , such that the process  $\tilde{X}_{n,k}(t)$  has left continuous sample paths and is  $\mathcal{A}_t$ -adapted. This is possible because of the martingale property of the  $E_\theta\{Z_m^{(k)}|\mathcal{A}_t\}$ ,  $t_{m-1}^{(k)} < t \leq t_m^{(k)}$ , and hence  $\tilde{X}_{n,k}(t)$  is  $\mathcal{A}_t$ -predictable.

From  $|\tilde{X}_{n,k}(t) - E_\theta\{X_n(t)|\mathcal{A}_t\}| \leq E_\theta\{|X_{n,k}(t) - X_n(t)||\mathcal{A}_t\}$  and the fact that  $X_{n,k} \rightarrow X_n$  in  $L^1$  as  $k \rightarrow \infty$ , we note that the sequence  $\tilde{X}_{n,k}$ ,  $k \geq 1$  is a Cauchy sequence in  $L^1$  and converges to some  $\mathcal{A}_t$ -predictable process  $\tilde{X}_n$  that equals  $E_\theta\{X_n(t)|\mathcal{A}_t\}$  almost surely for almost all  $t$ .

From the definition of  $X_n(t)$  and the fact that  $\{t \leq \tau_n^1\} \in \mathcal{A}_t$ , we have, almost surely,  $1_{\{t \leq \tau_n^1\}} E_\theta\{X_m(t)|\mathcal{A}_t\} = E_\theta\{X_n(t)|\mathcal{A}_t\}$  for  $m \geq n$  and hence there is an  $\mathcal{A}_t$ -predictable process  $\tilde{X}(t)$  such that

$$1_{\{\tau_n^1 > t\}} \tilde{X}(t) = \tilde{X}_n(t) = E_\theta[1_{\{\tau_n^1 > t\}} \lambda_{\theta,j}(t) | \mathcal{A}_t]$$

for almost all  $t$ , almost surely.

We now show that  $\tilde{X}(t)$  is  $\lambda_{\theta,j}(t)$ . For this, let  $A \in \mathcal{A}_t$ . Then the right-hand side of Eq. 2 equals

$$E_\theta \left[ \int_t^s 1_A X_n(u) du \right] = E_\theta \left[ 1_A \int_t^{s \wedge \tau_n^1} \tilde{X}(u) du \right].$$

By Eq. 2 the process  $N(t) - \int_0^t \tilde{X}(u) du$ ,  $t \geq 0$  is a local martingale, which gives the desired result.

**COROLLARY.** Suppose that  $\lambda_{\theta,j}(t)$  is integrable for almost all  $t$ . Then  $\lambda_{\theta,j}(t) = E_\theta\{\lambda_{\theta,j}(t)|\mathcal{A}_t\}$  almost surely, for almost all  $t$ .

Although the above result provides a means of computing  $\lambda_\theta(t)$  in terms of  $\lambda_{\theta,j}(t)$ , the actual computation is not easy. Instead, we approximate  $\lambda_\theta(t)$  by some  $\hat{\lambda}_\theta(t)$  that depends only on  $N(s)$ ,  $0 \leq s \leq t$ . The approximate log likelihood function is then

$$L_T(\theta) = \int_0^T \left\{ \sum_{j=1}^r \log \hat{\lambda}_{\theta,j}(t) dN_j(t) - \hat{\lambda}_{\theta,j}(t) dt \right\}. \quad [5]$$

Since  $T^{-1} \tilde{L}_T(\theta)$  depends essentially on the values of  $\hat{\lambda}_\theta(t)$  for large  $t$ , if  $T$  is large, we would expect  $L_T$  to be a good approximation to  $\tilde{L}_T$  for large  $T$  if  $\hat{\lambda}_\theta(t)$  is a good approximation to  $\lambda_\theta(t)$  for large  $t$ . By the Corollary of Theorem 1, and assuming that  $N(t)$  is a process with stationary increments,

we have that  $E\|\hat{\lambda}_\theta(t) - \lambda_\theta(t)\| \rightarrow 0$  as  $t \rightarrow \infty$ , which suggests that  $L_T$  is a good approximation to  $\tilde{L}_T$  for large  $T$  if  $\hat{\lambda}_\theta(t)$  is a good approximation to  $\lambda_\theta(t)$  for large  $t$ .

To illustrate our point, consider the self-exciting process when  $\lambda_\theta(t)$  is given by Eq. 3. In order that  $\lambda_{\theta,j}(t) \geq 0$  for all  $t$ , we shall assume that  $\alpha_{\theta,j} \geq 0$ ,  $g_{\theta,j}(t) \geq 0$  for all  $t$ . If the  $g_{\theta,j}$  are integrable and  $N(t)$  is of stationary increments with  $EN_j(1) = \mu_{\theta,j} < \infty$ , then the integral in Eq. 3 exists almost surely. Indeed

$$E_\theta \left\{ \int_{-\infty}^t g_{\theta,j}(t-s) dN_j(s) \right\} \leq \mu_{\theta,j} \int_{-\infty}^{\infty} g_{\theta,j}(t) dt < +\infty,$$

and hence the integral in the above left-hand side is finite almost surely. As an approximation of  $\lambda_\theta(t)$ , one might consider

$$\hat{\lambda}_\theta(t) = \alpha_\theta + \int_0^t g_\theta(t-s) dN(s),$$

which is evidently  $\mathcal{A}_t$ -measurable and would be a good approximation to  $\lambda_\theta(t)$  for large  $t$  if  $g_\theta(t) \rightarrow 0$  sufficiently fast.

#### Asymptotic Properties of the Approximate Log Likelihood

From now on, we shall suppose that the  $N(t)$  process is of stationary increments and metrically transitive in the sense of Doob (p. 510 of ref. 7). Let  $\bar{\Omega}$  be the space of  $\bar{\omega} = (\bar{\omega}_1, \dots, \bar{\omega}_r)$ , where the  $\bar{\omega}_i$  are nondecreasing integer-valued functions on  $(-\infty, \infty)$ ,  $\bar{\mathcal{A}}$  is the  $\sigma$ -field generated by the projections  $\pi_{st}$ :  $\bar{\omega} \rightarrow \bar{\omega}(t) - \bar{\omega}(s)$ ,  $s < t$ ,  $\bar{P}_\theta$  is the restriction to  $\bar{\mathcal{A}}$  of the image of  $P_\theta$  by the application  $N: \omega \rightarrow N(\cdot)$ , and  $T_h$  is the shift operator  $(T_h \bar{\omega})(t) = \bar{\omega}(t+h)$ . The stationarity means that  $T_h$  conserves the probability  $\bar{P}_\theta$ ; that is,  $\bar{P}_\theta(T_h^{-1}A) = \bar{P}_\theta(A)$ ,  $\forall A \in \bar{\mathcal{A}}$ , and metric transitivity means that the invariant sets, which are those sets  $A \in \bar{\mathcal{A}}$  with  $T_h^{-1}A = A$ , a.s., have probabilities zero or one. Now, from the fact that  $N(t) - N(s) = \pi_{st} \circ N$  and that  $T_h$  conserves the probability  $\bar{P}_\theta$ , one can show that  $\lambda_\theta(t) = \bar{\lambda}_\theta(t) \circ N$  with  $\bar{\lambda}_\theta(t+h) = \bar{\lambda}_\theta(t) \circ T_h$ . Hence if  $\lambda_\theta(t)$  is integrable, then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \lambda_\theta(t) dt = E\{\lambda_\theta(0)\}, \text{ almost surely.}$$

We assume that  $\theta \in R^k$  and  $\lambda_\theta(t)$ ,  $\hat{\lambda}_\theta(t)$  are twice continuously differentiable with respect to  $\theta$ , almost surely. We shall use the notation  $x_\theta^{(1)}$ ,  $x_\theta^{(2)}$  to denote the vector and the matrix of first and second derivatives of  $x_\theta$  with respect to  $\theta$ . Let  $\psi(t)$  be a function of the  $\lambda_{\theta,j}(t)$ ,  $\lambda_{\theta,j}^{(i)}(t)$ ,  $i = 1, 2$  such that  $\psi(t)$  is integrable. Then, by the same argument as above

$$\frac{1}{T} \int_0^T \psi(t) dt \rightarrow E\psi(0), \quad \text{a.s. } (T \rightarrow \infty). \quad [6]$$

In this section, we are interested in the limiting behavior of  $L_T(\theta)$ ,  $L_T^{(i)}(\theta^*)$ ,  $i = 1, 2$  as  $T \rightarrow \infty$ ,  $\theta^*$  being the true value of  $\theta$ . We have that  $dm_j = dN_j(t) - \lambda_j(t)dt$  when  $\theta = \theta^*$ , and putting  $\hat{\phi}_{\theta,j}(t) = \log \hat{\lambda}_{\theta,j}(t)$ , it follows that

$$L_T(\theta) = \sum_{j=1}^r \left[ \int_0^T \hat{\phi}_{\theta,j}(t) dm_j(t) + \int_0^T \{\hat{\phi}_{\theta,j}(t) \lambda_j(t) - \hat{\lambda}_j(t)\} dt \right] \quad [7]$$

$$L_T^{(1)}(\theta) = \sum_{j=1}^r \left[ \int_0^T \hat{\phi}_{\theta,j}^{(1)}(t) \{ \lambda_j(t) - \hat{\lambda}_{\theta,j}(t) \} dt + \int_0^T \hat{\phi}_{\theta,j}^{(1)}(t) dm_j(t) \right] \quad [8]$$

and

$$L_T^{(2)}(\theta) = \sum_{j=1}^r \left[ \int_0^T \hat{\phi}_{\theta,j}^{(2)}(t) \{ \lambda_j(t) - \hat{\lambda}_{\theta,j}(t) \} dt + \int_0^T \hat{\phi}_{\theta,j}^{(2)}(t) dm_j(t) - \int_0^T \hat{\phi}_{\theta,j}^{(1)}(t) \hat{\phi}_{\theta,j}^{(1)}(t)' \hat{\phi}_{\theta,j}(t) dt \right]. \quad [9]$$

To use the result 6 we replace  $\hat{\lambda}_{\theta,j}(t)$ ,  $\hat{\phi}_{\theta,j}(t)$ , ... by  $\lambda_{\theta,j}(t)$ ,  $\phi_{\theta,j}(t) = \log \lambda_{\theta,j}(t)$ , ... The following result can easily be proven and is quite useful.

LEMMA 1. Let  $\varepsilon_t, x_t$  be such that  $\varepsilon_t/x_t \rightarrow 0$  as  $t \rightarrow \infty$ . Also let  $\int_0^T |\varepsilon_t| d\mu(t) < \infty$  and  $\mu(T)^{-1} \int_0^T |x_t| d\mu(t)$  be bounded where  $\mu$  is a nondecreasing function with  $\mu(\infty) = \infty$ . Then  $\mu(T)^{-1} \int_0^T \varepsilon_t d\mu(t) \rightarrow 0$  as  $T \rightarrow \infty$ .

To obtain convergence results for the stochastic integrals with respect to  $dm_j(t)$  in Eqs. 7 and 8, we will need the following result, which is also of independent interest.

LEMMA 2. Let  $M_t, t \geq 0$  be a locally square integrable martingale with continuous natural increasing process  $\langle M \rangle_t$ . Let  $g_t, t > 0$  be a nondecreasing left continuous function with  $g_\infty = \infty$ , such that  $\langle M \rangle_t = O(g_t^{2-\varepsilon})$ ,  $\varepsilon > 0$ , almost surely as  $t \rightarrow \infty$ . Then  $M_T/g_T \rightarrow 0$  almost surely as  $T \rightarrow \infty$ .

Proof: Let  $a > 0$  be such that  $g_a > 0$ . Let  $c$  be an arbitrary positive constant. Define the stopping time  $\tau = \inf \{t: \int_a^t g_s^{-2} d\langle M \rangle_s \geq c\}$ . Set  $M_t = M_{t \wedge \tau}$ , and  $Z_t = \int_a^t g_s^{-1} d\tilde{M}_s = \int_a^{t \wedge \tau} g_s^{-1} dM_s$ . The process  $Z_t, t \geq a$  is a martingale with a natural increasing process  $\langle Z \rangle_t = \int_a^{t \wedge \tau} g_s^{-2} d\langle M \rangle_s \leq c, t \geq a$ . Since  $EZ_t^2 \leq c, t \geq a$ , we know (pp. 354 and 316 of ref. 7) that almost surely  $Z_{a+}, Z_{t-}, Z_{t+}, t \geq a, Z_\infty$  exist and  $Z_t$  is bounded on any finite interval. Set  $t_i^{(k)} = a + (T-a)i/k$  and write

$$g_T Z_T = \sum_{i=1}^k [g_{t_i^{(k)}} - g_{t_{i-1}^{(k)}}] Z_{t_i^{(k)}} + g_{t_{k-1}^{(k)}} [Z_{t_k^{(k)}} - Z_{t_{k-1}^{(k)}}] \\ = \int_0^T Z_t^{(k)} dg_t + \int_a^T g_t^{(k)} dZ_t,$$

where  $Z_t^{(k)} = Z_{t_i^{(k)}} = g_{t_i^{(k)}}^{-1}$  for  $t_{i-1}^{(k)} \leq t < t_i^{(k)}$ .

By the Lebesgue-dominated convergence theorem and the property of stochastic integral, almost surely as  $k \rightarrow \infty$

$$\int_a^{T-} Z_t^{(k)} dg_t \rightarrow \int_a^{T-} Z_{t+} dg_t$$

and

$$\int_a^T g_t^{(k)} dZ_t \rightarrow \int_a^T g_t dZ_t = \int_a^T d\tilde{M}_t = \tilde{M}_T - \tilde{M}_a.$$

Therefore,

$$g_T^{-1}(\tilde{M}_T - \tilde{M}_a) = Z_T - (1 - g_a g_T^{-1}) Z_\infty \\ - g_T^{-1} \int_a^{T-} (Z_{t+} - Z_\infty) dg_t.$$

Since  $g_T \uparrow \infty$  as  $T \rightarrow \infty$ , from Lemma 1, we obtain  $g_T^{-1} \tilde{M}_T \rightarrow 0$  almost surely as  $T \rightarrow \infty$ .

Now, we have  $\tilde{M}_t = M_t$  for all  $t$  on the set  $\{\tau = \infty\} = \{\int_a^\infty g_s^{-2} d\langle M \rangle_s \leq c\}$ . Since  $c$  is arbitrary, we obtain  $g_T^{-1} \tilde{M}_T \rightarrow$

0 almost surely as  $T \rightarrow \infty$  on the set  $\{\int_a^\infty g_s^{-2} d\langle M \rangle_s < \infty\}$ . By assumption  $\langle M \rangle_s \leq \text{const } g_s^{2-\varepsilon}$ , for all  $s \geq b$ , for some  $b \geq a$ , almost surely. So,

$$\int_b^\infty g_s^{-2} d\langle M \rangle_s < \infty; \text{ also } \int_a^\infty g_s^{-2} d\langle M \rangle_s < \infty.$$

The proof follows.

THEOREM 2. Suppose that

(i)  $P\{\lambda_{\theta,j}(t) \geq c\} = 0$  for some  $c > 0$  ( $c$  is independent of  $t$  by stationarity) and  $\lambda_{\theta,j}(t) - \lambda_{\theta,j}(t), \hat{\lambda}_{\theta,j}^{(i)}(t) - \lambda_{\theta,j}^{(i)}(t), i = 1, 2$  tend to zero almost surely as  $t \rightarrow \infty$ .

(ii)  $\lambda_{\theta,j}(t)$  is integrable, and  $\phi_{\theta,j}(t)$  and the elements of  $\phi_{\theta,j}^{(1)}(t), \phi_{\theta,j}^{(2)}(t)$  are square integrable with respect to the measure  $\lambda_j(t) dP$ .

Then, almost surely as  $T \rightarrow \infty$ ,

$$T^{-1} L_T(\theta) \rightarrow E \left[ \sum_{j=1}^r \phi_{\theta,j}(0) \lambda_j(0) - \lambda_{\theta,j}(0) \right] = \Lambda(\theta)$$

$$T^{-1} L_T^{(1)}(\theta^*) \rightarrow 0$$

$$T^{-1} L_T^{(2)}(\theta^*) \rightarrow -E \left\{ \sum_{j=1}^r \phi_j^{(1)}(0) \phi_j^{(1)}(0)' \phi_j(0) \right\} = -J.$$

Proof: Consider the right-hand side of Eq. 7. By (i),

$$\hat{\phi}_{\theta,j}(t) \lambda_j(t) - \hat{\lambda}_{\theta,j}(t) = \phi_{\theta,j}(t) \lambda_{\theta,j}(t) - \lambda_{\theta,j}(t) + o(t)$$

where in this proof  $o(t)$  denotes a quantity tending to zero almost surely as  $t \rightarrow \infty$ . By Lemma 1 and the fact that  $\phi_{\theta,j}(t) \lambda_j(t)$  is integrable, we obtain from Eq. 6 that  $T^{-1}$  times the last term of the extreme right-hand side of Eq. 7 tends to  $\Lambda(\theta)$  almost surely as  $T \rightarrow \infty$ . On the other hand,  $m_j(t)$  is a martingale with natural increasing process given by  $d\langle m_j \rangle_t = \lambda_j(t) dt$ . Therefore,  $M_{j,t} = \int_0^t \phi_{\theta,j}(t) dm_j(t)$  is also a martingale with natural increasing process  $\langle M_j \rangle_t = \int_0^t \phi_{\theta,j}^2(t) \lambda_j(t) dt$ . If  $\langle M_j \rangle_t = 0(t)$  almost surely as  $t \rightarrow \infty$ , then, by Lemma 2,  $M_{j,T}/T \rightarrow 0$  and hence  $T^{-1} L_T(\theta) \rightarrow \Lambda(\theta)$  almost surely as  $T \rightarrow \infty$ . Since from (i)

$$\hat{\phi}_{\theta,j}^2(t) \lambda_j(t) = \phi_{\theta,j}^2(t) \lambda_j(t) + \phi_{\theta,j}(t) \lambda_j(t) o(t) + \lambda_j(t) o(t)$$

and so  $M_{j,T}/T \rightarrow E\{\phi_{\theta,j}^2(0) \lambda_j(0)\}$  almost surely as  $T \rightarrow \infty$ , we obtain the result.

The proofs for the convergence of  $T^{-1} L_T^{(1)}(\theta^*)$  and  $T^{-1} L_T^{(2)}(\theta^*)$  follow analogously.

Consider now the asymptotic distribution of  $T^{-1/2} L_T(\theta^*)$ . From Eq. 8, we expect that this is asymptotically distributed like

$$Z_T = T^{-1/2} \int_0^T \sum_{j=1}^r \phi_j^{(1)}(t) dm_j(t). \quad [10]$$

The asymptotic distribution of  $Z_T$  can be obtained from the following result, which is of independent interest.

LEMMA 3. Let  $M_t, \mathcal{A}_t, t \geq 0$  be a square integrable martingale with natural increasing process  $\langle M \rangle_t$  satisfying  $d\langle M \rangle_t = X_t dt$ . Suppose that there is a semigroup of shift operator  $T_s, s \geq 0$ , conserving the probability and metrically transitive, such that  $T_h^{-1} \mathcal{A}_t = \mathcal{A}_{t+h}, (M_t - M_s) \circ T_h = M_{t+h} - M_{s+h}$ . Then as  $T \rightarrow \infty, T^{-1/2} M_T$  is asymptotically normal with mean zero and variance  $E X_0$ .

Proof: Let  $n_T$  be integers such that  $\Delta_T = T/n_T \rightarrow 0$  as  $T \rightarrow \infty$ . Set  $Y_{T,j} = T^{-1/2} (M_{j\Delta_T} - M_{(j-1)\Delta_T})$ . Then, as  $T \rightarrow \infty, T^{-1/2} M_T = T^{-1/2} M_0 + \sum_{j=1}^{n_T} Y_{T,j}$  is asymptotically normal

with mean zero and variance  $EX_0$ , provided

- (i)  $\sum_{j=1}^{n_T} E\{Y_{T,j}^2 | \mathcal{A}_{(j-1)\Delta_T}\} \rightarrow EX_0$  in probability  
 (ii)  $\sum_{j=1}^{n_T} E[Y_{T,j}^2 1_{\{Y_{T,j}^2 > \varepsilon\}}] \rightarrow 0, \varepsilon > 0.$

To verify *i*, observe that the sum in *i* equals

$$\frac{1}{T} \int_0^T X_t^{(T)} dt = \langle M \rangle_T / T + \frac{1}{T} \int_0^T (X_t^{(T)} - X_t) dt,$$

where  $X_t^{(T)} = E(X_t | \mathcal{A}_{j\Delta_T})$  for  $j\Delta_T \leq t < (j+1)\Delta_T$ . Clearly  $X_{t+h} = X_t \circ T_h$  and  $E(X_{t+h} | \mathcal{A}_{s+h}) = E(X_t | \mathcal{A}_s) \circ T_h$ . Therefore, as  $\Delta_T \rightarrow 0$ ,

$$\sup_{0 \leq t \leq T} E|X_t^{(T)} - X_t| \leq \sup_{0 \leq u \leq \Delta_T} E|E(X_1 | \mathcal{A}_{1-u}) - X_1| \rightarrow 0.$$

Note that  $X_t$  is  $\mathcal{A}_t$ -measurable because this process is predictable. Thus, the sum in *i* differs from  $\langle M \rangle_T / T$  by a term tending to 0 in the mean and since  $\langle M \rangle_T / T \rightarrow EX_0$  by the metric transitivity of  $T_h$ , we obtain the result.

To verify *ii*, observe that the  $Y_{T,j}, j = 1, \dots, n_T$  have the same distribution. Putting  $\xi(t) = M_t^2 - M_0$ , the result will follow by verifying that

$$n_T[E(Y_{T,1}^2 1_{\{Y_{T,1}^2 > \varepsilon\}})] = \Delta_T^{-1} E[\xi(\Delta_T) 1_{\{\xi(\Delta_T) > \varepsilon T\}}]$$

tends to zero as  $T \rightarrow \infty$ .

We now show that the difference between  $T^{1/2}L_T(\theta^*)$  and  $Z_T$ , defined by Eqs. 7 and 10, respectively, tends to zero in probability as  $T \rightarrow \infty$ .

**LEMMA 4.** Let  $M_t, t \geq 0$  be a locally integrable martingale with continuous natural increasing process  $\langle M \rangle_t$ . Let  $h_t \geq 0$ . If  $\langle M \rangle_T / h_T^2 \rightarrow 0$  in probability as  $T \rightarrow \infty$ , then so does  $M_T / h_T$ .

*Proof:* Let  $\varepsilon > 0$ . Define the stopping time  $\sigma_\varepsilon$  to be the value of  $s$  such that  $\langle M \rangle_s = \varepsilon h_s^2$ . Then as  $T \rightarrow \infty$ ,  $P\{M_T \wedge \sigma_\varepsilon \neq M_T\} \rightarrow 0$ . On the other hand,  $E(M_T^2 \wedge \sigma_\varepsilon) = E\langle M \rangle_{T \wedge \sigma_\varepsilon} \leq E\langle M \rangle_{\sigma_\varepsilon} = \varepsilon h_{\sigma_\varepsilon}^2$  and, hence, by the Chebyshev inequality,

$$P\{|M_T / h_T| > \delta\} \leq P\{M_T \wedge \sigma_\varepsilon \neq M_T\} + \varepsilon / \delta^2,$$

from which the result follows.

**THEOREM 3.** Suppose the conditions of Theorem 2 are satisfied and also

$$(i) T^{-1/2} \int_0^T \hat{\phi}_j^{(1)}(t) \{\hat{\lambda}_j(t) - \lambda_j(t)\} dt \rightarrow 0$$

in probability as  $T \rightarrow \infty$ . Then  $T^{-1/2}L_T(\theta^*)$  is asymptotically normal with mean zero and covariance matrix  $J$  given in Theorem 2.

*Proof:* Let  $\alpha$  be a vector and  $Z_T$  be given by Eq. 10. Then  $M_T = T^{1/2}\alpha'Z_T, T \geq 0$  is a martingale with natural increasing process

$$\langle M \rangle_t = \int_0^t \sum_{j=1}^r \{\alpha' \phi_j^{(1)}(s)\}^2 \lambda_j(s) ds.$$

By Lemma 3,  $\alpha'Z_T$  is asymptotically normal with mean zero and variance  $\alpha'J\alpha$ . On the other hand, by *i* and Lemma 4 with  $dM_t = \sum_{j=1}^r \{\phi_j^{(1)}(t) - \hat{\phi}_j^{(1)}(t)\} dm_j(t), h_t = \sqrt{t}$ , we see that  $T^{-1/2}L_T(\theta^*) - Z_T \rightarrow 0$  in probability as  $T \rightarrow \infty$ . The result follows.

## Asymptotic Properties of the Maximum Likelihood Estimator

We now study the asymptotic properties of the estimate  $\theta_T$ , which maximizes  $L_T$  in  $\Theta$ . For convenient reference, we state the general result.

**THEOREM 4. 1.** Let  $\Lambda_T$  be a random function on  $\theta \in \mathbb{R}^k$ , satisfying

- (i)  $\Lambda_T(\theta) \rightarrow \Lambda(\theta)$  almost surely as  $T \rightarrow \infty$ , with  $\Lambda$  being continuous, admitting a unique maximum  $\theta^*$ ,  
 (ii) For any  $\theta \neq \theta^*$ ,

$$\lim_{T \rightarrow \infty} \sup_{\theta' \in U(\theta)} \{\Lambda_T(\theta') - \Lambda_T(\theta)\} \rightarrow 0 \text{ a.s.}$$

as the neighborhood  $U(\theta)$  of  $\theta$  shrinks to  $\theta$ .

Then any  $\theta_T$  realizing the maximum of  $\Lambda_T$  on a compact  $C$  of  $\Theta$  containing  $\theta^*$ , converges almost surely to  $\theta^*$  as  $T \rightarrow \infty$ .

2. Suppose that  $\theta^*$  is an interior point of  $\Theta$ , and  $\Lambda$  admits continuous first and second derivatives with respect to  $\theta$ , denoted by the vector  $\Lambda_T^{(1)}$  and the matrix  $\Lambda_T^{(2)}$ , satisfying

(iii) As  $T \rightarrow \infty$ ,  $\Lambda_T^{(2)}(\theta^*) \rightarrow -J$  in probability, and  $\sqrt{T}\Lambda_T^{(1)}(\theta)$  is asymptotically normal with mean zero and covariance matrix  $J$ .

(iv) For every  $\varepsilon > 0$ ,

$$\liminf_{T \rightarrow \infty} P\{|\Lambda_T^{(2)}(\theta) - \Lambda_T^{(2)}(\theta^*)| < \varepsilon \quad \forall \theta \in U(\theta^*)\}$$

increases to 1 as the neighborhood  $U(\theta^*)$  of  $\theta^*$  shrinks to  $\theta^*$ .

Then  $\theta_T$  is asymptotically normal with mean  $\theta^*$  and covariance matrix  $T^{-1}J^{-1}$ . Moreover, if  $\theta_T$  is  $T^{1/2}$ -consistent, that is the distributions of  $\sqrt{T}(\theta_T - \theta^*)$  are tight, then  $\theta_T - \{\theta_T - \Lambda_T^{(2)}(\theta_T)^{-1}\Lambda_T^{(1)}(\theta_T)\} \rightarrow 0$  in probability as  $T \rightarrow \infty$ .

We apply the above results with  $\Lambda_T = T^{-1}L_T$ . By Theorem 2, condition *i* is satisfied except for the continuity of  $\Lambda(\theta)$ , which we shall assume. To see that  $\theta^*$  realizes the maximum of  $\Lambda(\theta)$ , write

$$\Lambda(\theta) - \Lambda(\theta^*) =$$

$$\sum_{j=1}^r [E \log \{\lambda_{\theta,j}(0) \lambda_j^{-1}(0)\} \lambda_j(0) + \lambda_j(0) - \lambda_{\theta,j}(0)]$$

and note that  $\log x \leq x - 1$  with equality if and only if  $x = 1$ , we get  $\Lambda(\theta) \leq \Lambda(\theta^*)$  with equality if and only if  $\lambda_{\theta,j}(0) = \lambda_j(0)$  almost surely, implying, by stationarity that  $\lambda_{\theta,j}(t) = \lambda_j(t)$  almost surely. If the parametrization is such that for  $\theta' \neq \theta$ ,  $\lambda_{\theta',j}(t)$  is not equal to  $\lambda_{\theta,j}(t)$  for all  $t$  almost surely, then  $\theta^*$  is the unique maximum of  $\Lambda$ .

Theorems 2 and 3 show that condition *iii* is satisfied. So all we need is to verify conditions *ii* and *iv*. This would require rather strong assumptions on  $\lambda_\theta(t)$  and  $\lambda_\theta(t)$ . A sufficient set of assumptions is as follows.

A0: For any compact  $C$  of  $\Theta$ , there is a  $c > 0$  such that  $\lambda_{\theta,j}(t) \geq c$  almost surely for all  $t$ , all  $\theta \in C$ .

A1: For any compact  $C$  of  $\Theta$

$$\lim_{t \rightarrow \infty} \sup_{\theta \in C} |\hat{\lambda}_{\theta,j}(t) - \lambda_{\theta,j}(t)| = 0 \text{ a.s.,}$$

$$E\left\{\sup_{\theta \in C} \phi_{\theta,j}^2(t)\right\} \lambda_j(t) < \infty.$$

A2: For some compact neighborhood  $U$  of  $\theta^*$ ,

$$\lim_{t \rightarrow \infty} \sup_{\theta \in U} \|\hat{\lambda}_{\theta,j}^{(i)}(t) - \lambda_{\theta,j}^{(i)}(t)\| = 0 \text{ a.s.,} \quad i = 1, 2,$$

$$E \left[ \left\{ \sup_{\theta \in U} \|\phi_{\theta,j}^{(2)}(t)\|^2 \right\} \lambda_j(t) \right] < \infty,$$

$$E \left\{ \sup_{\theta \in U} \|\lambda_{\theta,j}^{(2)}(t)\| \right\} < \infty.$$

**THEOREM 5.** *Under assumptions A0 and A1, condition ii of Theorem 4 is satisfied, and under assumptions A0 and A1, condition iv of Theorem 4 is satisfied.*

*Proof:* Let  $U(\theta)$  be a compact neighborhood of  $\theta$ . Then from Eq. 7,

$$\sup_{\theta' \in U(\theta)} T^{-1} L_T(\theta') \leq T^{-1} \sum_{j=1}^r \left[ \int_0^T \left\{ \sup_{\theta' \in U(\theta)} \hat{\phi}_{\theta',j}(t) \right\} dN_j(t) - \int_0^T \left\{ \inf_{\theta' \in U(\theta)} \hat{\lambda}_{\theta,j}(t) \right\} dt \right].$$

By the same argument as in the proof of Theorem 2, the above right-hand side is seen to converge almost surely as  $T \rightarrow \infty$  to

$$\sum_{j=1}^r \left[ E \left\{ \left\{ \sup_{\theta' \in U(\theta)} \phi_{\theta',j}(t) \right\} \lambda_j(t) \right\} - E \left\{ \inf_{\theta' \in U(\theta)} \lambda_{\theta,j}(t) \right\} \right].$$

By the monotone convergence theorem, as  $U(\theta)$  shrinks to  $\theta$

the above expression converges to  $\Lambda(\theta)$  and hence

$$\lim_{T \rightarrow \infty} \sup \left\{ \sup_{\theta' \in U(\theta)} T^{-1} L_T(\theta') \right\} \rightarrow \Lambda(\theta) \text{ a.s.},$$

which gives the result.

Verification of condition iv follows essentially analogously.

We are highly obliged to Prof. R. N. Bhattacharya for the very careful and critical examination of the first draft of this paper. His extremely valuable comments and helpful suggestions are gratefully acknowledged. The research of M.L.P. was supported in part by Office of Naval Research Contract N00014-85-K-0648.

1. Jacobsen, M. (1982) *Statistical Analysis of Counting Process*, Lecture Notes in Statistics (Springer, New York), Vol. 12, p. 63.
2. Jacod, J. (1975) *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **31**, 235–253.
3. Lipster, R. S. & Shirayev, A. N. (1978) *Statistics of Random Processes II* (Springer, Berlin), p. 252.
4. Hawkes, A. G. (1971) *Biometrika* **58**, 83–90.
5. Hawkes, A. G. & Oakes, D. (1974) *J. Appl. Probab.* **11**, 493–503.
6. Brémaud, P. & Jacod, J. (1977) *Adv. Appl. Probab.* **9**, 369–416.
7. Doob, J. L. (1953) *Stochastic Processes* (Wiley, New York).